

Objetivos en el desarrollo del Matrices-TAI

- Ofrecer una **estimación de la inteligencia basada en estímulos no verbales**, lo que permite utilizar el test incluso con personas que no conocen o dominan el español o con capacidades comunicativas reducidas (personas procedentes de otros países que no dominan el idioma, personas con problemas con el lenguaje, etc.).
- Crear un instrumento psicométrico cuyas puntuaciones sean **fiables y permitan realizar inferencias válidas** a través de un amplio rango de aptitud.
- Obtener una **prueba de fácil aplicación y corrección, con el máximo ahorro de tiempo**.

Para los objetivos anteriores se utilizan modelos de Teoría de la Respuesta al Ítem, que permiten la construcción de un **TEST ADAPTATIVO INFORMATIZADO (TAI)**. En un TAI se **ajusta la dificultad de la tarea a diferentes niveles aptitudinales** para aumentar la capacidad de discriminación del test y su utilidad en diferentes contextos y finalidades.

Elementos de un TAI

Banco de ítems tipo matrices

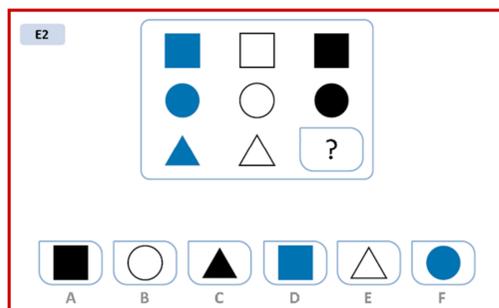


Figura 1.1. Ítem de ejemplo del Matrices

Se parte de un banco inicial de 326 ítems. Tras un estudio piloto se seleccionan 149 ítems que fueron aplicados a muestras de aptitud heterogénea de 12,280 personas, en un diseño de anclaje (149 ítems repartidos en seis formas de 52 ítems de dificultad variada).

Prácticamente todos los ítems ajustaron al modelo unidimensional y al modelo de Teoría de la Respuesta al Ítem

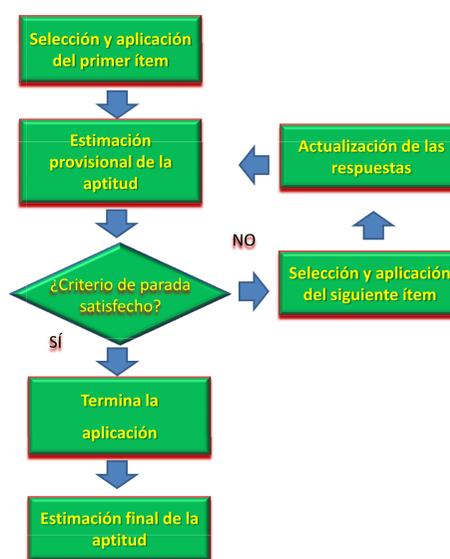
Diseño de anclaje

Forma	N.º de ítems compartidos entre las formas de tipificación					
	Nivel A	Nivel B	Nivel C	Nivel D	Nivel E	Nivel F
Nivel A	52	39	18	8	2	0
Nivel B	39	52	27	9	2	0
Nivel C	18	27	52	25	12	1
Nivel D	8	9	25	52	24	6
Nivel E	2	2	12	24	52	28
Nivel F	0	0	1	6	28	52

Resultados de ajuste

Forma	Ítems	N	χ^2	G.I.	RMSEA	CFI	TLI	SRMR	% varianza	Menor peso
Nivel A	52	467	1459,9	1.274	0,018	0,958	0,956	0,089	28	0,26
Nivel B	52	1528	2129,6	1.274	0,021	0,946	0,944	0,061	25	0,21
Nivel C	51	2362	3436,7	1.224	0,028	0,944	0,941	0,057	27	0,26
Nivel D	52	2178	2647,7	1.274	0,022	0,958	0,956	0,050	27	0,26
Nivel E	52	2235	2663,8	1.274	0,022	0,964	0,963	0,047	27	0,25
Nivel F	52	1881	2089,2	1.274	0,018	0,974	0,973	0,043	26	0,26

Algoritmo adaptativo



Procedimiento de arranque: Al comenzar la prueba no se dispone de ninguna información sobre el nivel de aptitud del evaluado, por tanto debe especificarse un procedimiento de arranque. Se selecciona un nivel de habilidad de una distribución normal con la media del grupo, truncada entre una desviación típica por debajo de ese nivel medio y una desviación típica por encima de ese nivel.

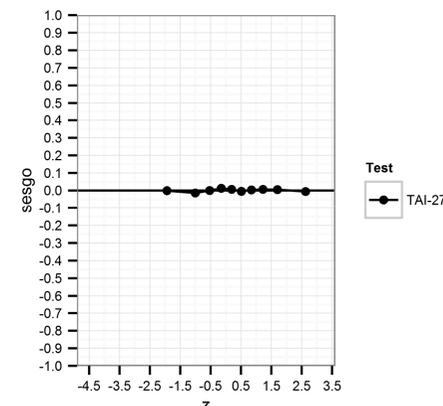
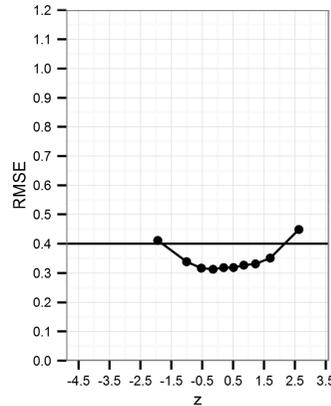
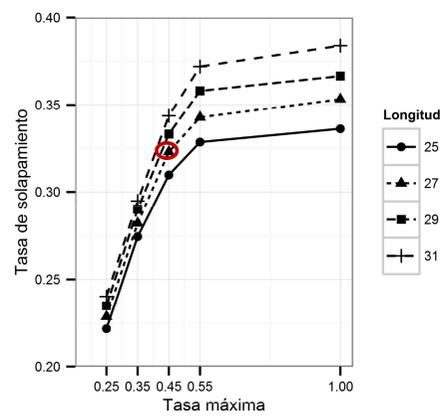
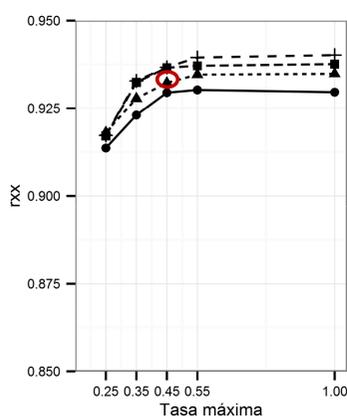
Estimación provisional y final del nivel de aptitud: Durante los primeros cinco ítems, existe poca información sobre el nivel de habilidad y se establece una estimación bayesiana EAP. Después del quinto ítem se utiliza el procedimiento de estimación de **Máxima Verosimilitud Ponderada (WLE; Warm, 1989; Wang y Hanson, 1999)**, que produce estimaciones más insesgadas que otros procedimientos (Kim y Nicewander, 1993).

Selección de ítems: El criterio de selección de ítems es el corazón de un TAI. En nuestro caso, se combinaron dos algoritmos de selección de ítems: a) Durante los cinco primeros ítems, se seleccionan los mejores ítems en base al criterio de **máxima proximidad** (i.e., se selecciona el ítem cuyo punto de máxima información está más próximo al nivel de rasgo provisional). Los ítems más discriminativos eran excluidos del bloque de ítems elegibles; b) Después del quinto ítem aplicado se aplica el método de selección de ítems de **máxima información**. En este caso, todos los ítems del banco eran elegibles.

Control de la exposición: La vida útil de un banco depende de los controles de seguridad de la prueba, especialmente en contextos de acreditación, promoción o selección. Se establecen dos mecanismos de control de la exposición: a) Tasa máxima de exposición del 45%, aplicando el **método de elegibilidad** de Van der Linden y Veldkamp (2004, 2007); b) Durante los cinco primeros ítems, se aplica el método **"Randomesque"** (Kingsbury y Zara, 1989), que consiste en seleccionar al azar entre los cinco mejores ítems.

Procedimiento de parada: Se establece como criterio de parada un número prefijado de 27 ítems para todos los sujetos

Propiedades del TAI



Se realizó un estudio de simulación para comprobar las características óptimas del TAI. Se consideraron cuatro criterios de parada (25, 27, 29 y 31 ítems) y cinco tasas máximas de exposición (0,25, 0,35, 0,45, 0,55 y 1). Como variables dependientes, se consideraron un indicador de fiabilidad (la correlación al cuadrado entre el nivel de rasgo estimado y el nivel de rasgo verdadero) y la tasa de solapamiento (que indica la proporción de ítems compartidos en promedio por dos evaluados al azar).

Los resultados indican que para todos los niveles contrastados se encuentran indicadores de fiabilidad excelentes (por encima de 0,9). El mayor incremento se produce al incrementar la tasa máxima de exposición de 0,25 a 0,35, mientras que flexibilizar de 0,45 a 0,55 apenas produce incrementos. En relación a la longitud del test, se encuentran incrementos muy pequeños al pasar de 27 a 31 ítems.

En relación a la tasa de solapamiento, se encuentran tasas de solapamiento por debajo del 40% para todas las condiciones.

Se obtuvieron dos indicadores de calidad de las estimaciones: (a) La Raíz del Error Cuadrático Medio (RMSE), que es la raíz del promedio de las diferencias al cuadrado entre el nivel de habilidad estimado y el nivel de habilidad verdadero; (b) El sesgo (un valor positivo indica que el nivel se sobrestima, un valor negativo que se subestima).

En relación al RMSE, se observa que, en general, la precisión del TAI es alta. Mediante el TAI se obtienen niveles de precisión homogéneos a lo largo de todos los niveles de rasgo. Para el TAI, en la mayor parte de los casos se obtienen RMSE por debajo de 0,4 (y siempre por debajo de 0,5). En ausencia de sesgo, RMSE de 0,3, 0,4 y 0,5 son equivalentes a fiabilidades de 0,91, 0,84, y 0,75 respectivamente. En relación a los niveles de sesgo, se encuentran resultados próximos a cero para todos los niveles.

Conclusiones

Puede concluirse que las puntuaciones obtenidas con el algoritmo del Matrices-TAI poseen ausencia de sesgo, así como una precisión adecuada y homogénea a lo largo del nivel de aptitud. Por tanto, el uso de este instrumento puede resultar útil en el contexto aplicado por la sencillez de su aplicación y su eficiencia.

REFERENCIAS:

- Kim, J. K. y Nicewander, W. A. (1993). Ability estimation for conventional tests. *Psychometrika*, 58(4), 587-599.
- Kingsbury, G. G. y Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2(4), 359-375.
- van der Linden, W. J. y Veldkamp, B.P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational & Behavioral Statistics*, 29, 273-291.
- van der Linden, W.J. y Veldkamp, B.P. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics*, 32, 398-418.
- Warm, T. A. (1989). Weighted Likelihood Estimation of ability in Item Response Theory. *Psychometrika*, 54(3), 427-450.
- Wang, T. y Hanson, B. A. (1999). Reducing Bias in TAI Trait Estimation: A Comparison of Approaches. *Applied Psychological Measurement*, 23(3), 263-278.